

# GUJARAT TECHNOLOGICAL UNIVERSITY

Chandkheda, Ahmedabad

Affiliated



## Gandhinagar Institute of Technology

A Report On-

### Plagiarism Analyzer

Under the subject of  
DESIGN ENGINEERING – 2B  
B.E.III, Semester – VI  
(Information Technology Branch)

Submitted by  
Group:

Sr.	Name of student	Enrolment No.
1.	Himanshu Parmar	200120116091
2.	Viral Dedaniya	200120116076
3.	Harsh Jagani	200120116060
4.	Naimish Pambhar	200120116112

Prof. Pooja J. Shah  
(Faculty Guide)

Dr. Nimisha Patel  
Head of the Department

Academic year  
(2023-2024)

## **Abstract**

The "Plagiarism Analyzer" provides functionality to check for plagiarism. It checks the text files for plagiarism and also checks the images if it is copied or not and generates the resultant image the difference image which shows the difference between the two images if the generated image is all black then the two images are identical.

The project works on the latest technology, it uses python 3.8 programming language. For the code editor, we use IntelliJ IDE, some modules are also required like difflib, cv2 and numpy.

# Contents

**Abstract**

**Contents**

**List of Figures**

**Chapter 1 Introduction**

1.1	About Domain/Area	5
1.2	About the Definition	5
1.3	Users	6
1.4	Modules and Functionalities	6

**Chapter 2 Literature Review**

2.1	Observation on Domain	7
2.2	Usage and Advantages	8
2.3	Literature review related to your definition	10

**Chapter 3 Design Consideration to detail design Part**

3.1	AEIOU	11
3.2	Empathy	13
3.3	Ideation Canvas	14
3.4	Product Development Canvas	15
3.5	Rough Prototype	16

**Chapter 4 Conclusion & Future Work**

4.1	Conclusion	17
-----	------------	----

<b>Appendix A</b>	Photos of Canvases	18
-------------------	--------------------	----

<b>References</b>		24
-------------------	--	----

## List of Figures

3.1	Plagiarism Image 1	9
3.2	Plagiarism Image 2	9
3.3	AEIOU Canvas	12
3.4	Empathy Canvas	13
3.5	Ideation Canvas	14
3.6	Product Development Canvas	15
3.7	Rough Prototype	16
3.8	Learning Need Matrix	17
3.9	Mind Map	18

# Chapter 1

## Introduction

### 1.1 About Domain/Area

#### (Plagiarism Finder, Image Difference)

Plagiarism detection or content similarity detection is the process of locating instances of plagiarism or copyright infringement within image or text.

### 1.2 About the Definition

The word plagiarism originates from the Latin word ‘plagiarius’.

The software system takes a submitted text as input, and compares the text against a set of publicly available and privately held documents, resulting in a similarity report.

The anti-plagiarism software enables the scanning of documents against all the documents sources, the software tool looks to match a string of eight to ten words.

### **1.3 Users**

Plagiarism Analyzer is used by limited people. Here are some common users:

- Teachers
- Students
- Educational Institutes
- Bloggers
- News Channels
- Media Companies
- Content Creators etc...

### **1.4 Modules and Functionalities**

#### **1.4.1 Functionalities**

- To provide facilities to improve the work and study culture
- To find plagiarism in text and images
- To find the difference between two images
- To generate an image to show the difference

## Chapter 2

### Literature Review

#### 2.1 Observation on Domain

In a world where information can be accessed from anywhere with a secure internet connection, developing research papers and articles may seem like a walk in the park. It's easy to think that you can find useful resources and information within seconds on the internet. However, in case you are thinking of copying and pasting already published work, then you should know that's considered stealing of intellectual property. It is often referred to as plagiarizing.

A plagiarism checker uses advanced database software to scan for matches between your text and existing texts. They are used by universities to scan student assignments. There are also commercial plagiarism checkers you can use to check your own work before submitting.

Behind the scenes, plagiarism checkers crawl web content and index it, scanning your text for similarities against a database of the existing content on the internet. Exact matches are highlighted using keyword analysis. Some checkers can also identify non-exact matches (paraphrasing plagiarism).

## Usage and Advantages

**Provides links to the source text** - The good thing with plagiarism checker is that they provide the links for the source text that is plagiarized. This makes it easier for anyone using the software to determine whether the results presented are based on substantial evidence. You get to compare the source text and what you have to find the duplicate content.

**Speed** - Plagiarism checkers work fast. These software applications can search the entire web within a matter of seconds. They then match the text in question with other documents from around the world and highlight any part/paragraph/section that has been plagiarized within a short time. The speed of such software is incredible, and you get to produce authentic content quickly.

**Provides the similarity percentage** - Whenever you write appropriate content and run it through a plagiarism checker, it provides you with the percentage of plagiarism. You get to know how authentic your content is and if it needs editing. Most universities use this plagiarism software to check papers so that the instructors can determine if their work is within the standard percentage rate acceptable to them. The similarity enables them to know if a particular article is copied or not.

**Useful for writing original pieces within a short time** - Plagiarism checkers have proved to be quite crucial, especially to web writers who are required to write authentic content within a quick turnaround. Usually, these guys face some tight deadlines, and as such, it can be challenging to get the time for checking the originality of the texts they write. However, with a plagiarism checker, it becomes easier to check the uniqueness of their web articles and blogs. The last thing you need as a writer is to submit your work only to find out later it was full of plagiarism. Not only does it make you lose credibility, but it also lowers your confidence in writing. This software provides an amicable solution, so you don't have to waste too much time regardless of the content.

**Helps in checking your paraphrasing prowess** - Plagiarism checkers only highlight content that is an exact match of the original author's words. As such, if you have paraphrased your material well enough, then it won't show any plagiarism. You get to know how excellent your paraphrasing skills are if you don't get any texts highlighted as having been copied.

**Helps in staying within the regulatory and ethical limits** - Plagiarism checkers play a significant role for both writers and academicians. For them, they have to operate within some regulatory and ethical limits if they want to excel in any field of education or work. For them, having content with plagiarism is one thing they have to avoid as much as possible. Everyone wants an authentic job since it shows your seriousness and intellectual level as well.





[3.1 Plagiarism Image 1]



[3.1 Plagiarism Image 1]

## **2.2 Problem Definition**

These are some problems if you don't deal with plagiarism

Plagiarism is academic misconduct in which students use someone's ideas or information in their work without proper referencing. While western culture insists that sources of words, ideas, images, and sounds be documented for academic purposes, "plagiarism is now recognized as a serious problem, especially in university where students are just copying words from websites or someone else's work"

Student plagiarism is problematic in all academic disciplines as a result of various factors and is a current social phenomenon that universities must address in order to preserve academic integrity. There are issues relating to defining plagiarism, course and assignment design to prevent plagiarism and best practices concerning plagiarism. However, while plagiarism may be related to the intention of cheating, it may also occur through the sharing of information with friends who have been given the same topic for an assignment and are meant to be working independently.

## Chapter 3 Design

### Consideration to detail design Part

#### 3.1 AEIOU

AEIOU,      A - stands for Activities  
                  E -stands for Environment  
                  I -stands for Interaction  
                  O -stands for Objects  
                  U -stands for User

This sheet deals with the normal survey our team carried out in order to observe the actual scenario at the crossroads which are having traffic issues and the situations people deal with. The observation can be anything respective as well as irrespective of the particular project.

#### Activity:

Activity is the condition in which things are happening or being done. So here in the AEIOU sheet, Activities are the ones that we observed going on at the location. We observed the activities done by the people present there and noted them.

- Upload original Image and doubtful image
- comparing the two images
- finding the difference
- Detect image plagiarism
- text plagiarism detection

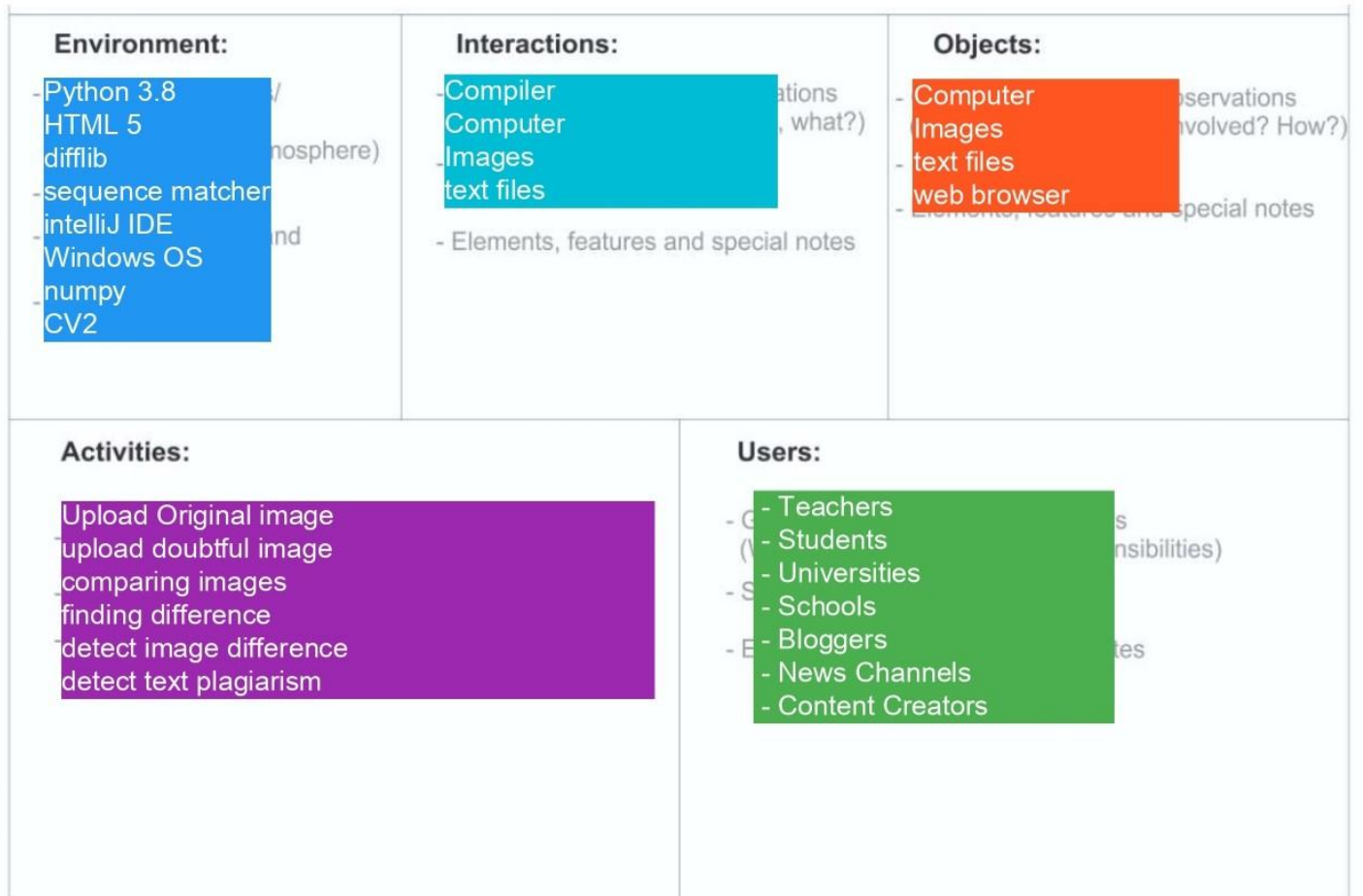
#### Environment:

In this, we observed the entire area of the location and noted down the places or things that surrounded the location. They are:

- Operating Systems like windows 10
- Python 3.8 or higher
- HTML 5
- diffliib
- sequence matcher
- IntelliJ IDE
- NumPy
- cv2

**Interaction: -**

- Compiler
- Computer
- Images
- Document Files



[3.1 AEIOU Canvas]

### 3.2 Empathy Canvas

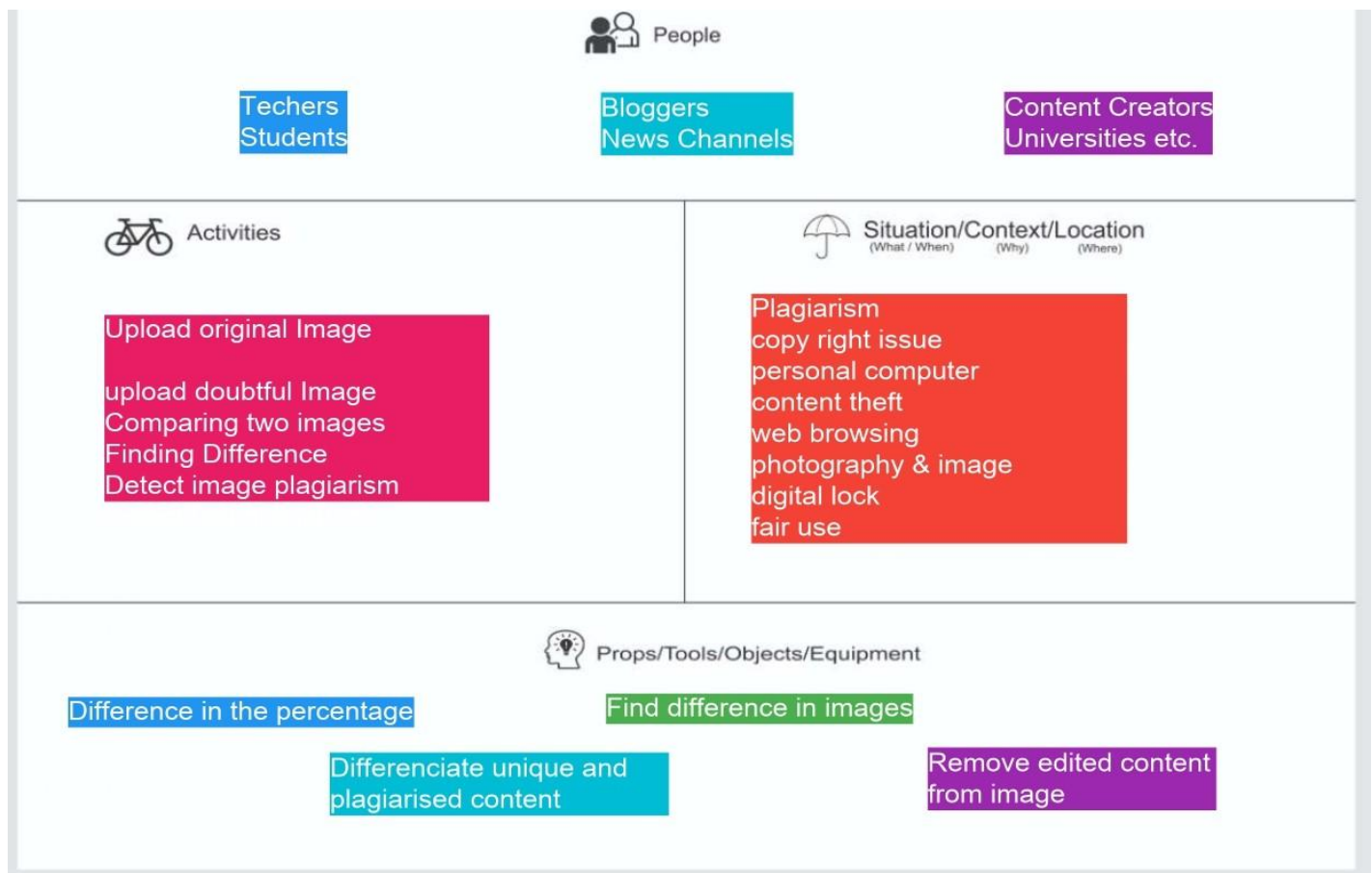
Understanding the problem of society is one of the biggest challenges for engineering students as they were making projects on imaginary ideas. Understanding the domains of the problem in broader sense emphasized on interacting with the people of your domain area which included more of casual talk rather than a technical sessions. It focused on making people comfortable in discussing their problem with our students. And this activity was named as Story Boarding/Empathy canvassing. Here we would like to explain the aspects of this canvas with the help of the example taken up by our team. We started to identify people (i.e., users and stakeholders) related to our domain directly or indirectly. Then we listed out activities which were involved in public services. As we need to focus on people's emotion, we wrote happy and sad story which helps us to get more attached with the users.

USER		STAKEHOLDERS
Teachers Students Bloggers	New Channels Content Creators	Developers
<b>ACTIVITIES</b> Upload Original image Upload doubtful image Comparing image Finding Difference Detect image plagiarism Detect Text plagiarism		
<b>STORY BOARDING</b> <b>HAPPY</b> I make Edits and NFT images, it gets stolen pretty quick which leads into loss with the help of it I can prevent it from getting stolen.		
<b>HAPPY</b> I occasionally upload blog on the internet and i have seen my content on someone else's website, I claimed it as an owner with the help of this I can finally prove its my original work.		
<b>SAD</b> It is very good project but it only detects the plagiarism and cant remove it.		
<b>SAD</b> It can not find if the video is plagiarised or not like youtube does because I am video editor and if it does that then it will be very helpful		

### 3.3 Ideation Canvas

An ideation canvas is a rough whiteboard where ideas can be stretched into any limits or dimensions. Ideation session is not aimed at finding solutions to the defined problem. But to define the best possible problem and stretch out its possible scope. The field is set, and the overall agenda is to build the clones of the ideas and pivot them throughout the canvas to discover new possibilities.

In ideation canvas we firstly started with people, where we simply thought about the people for whom we want to solve the problem. For our System we listed down the people, then listed out whatever activity every segment of people does. Then thought for context/location/situation and finally for possible solutions.











[3.3 Ideation Canvas]



### 3.4 Product Development Canvas

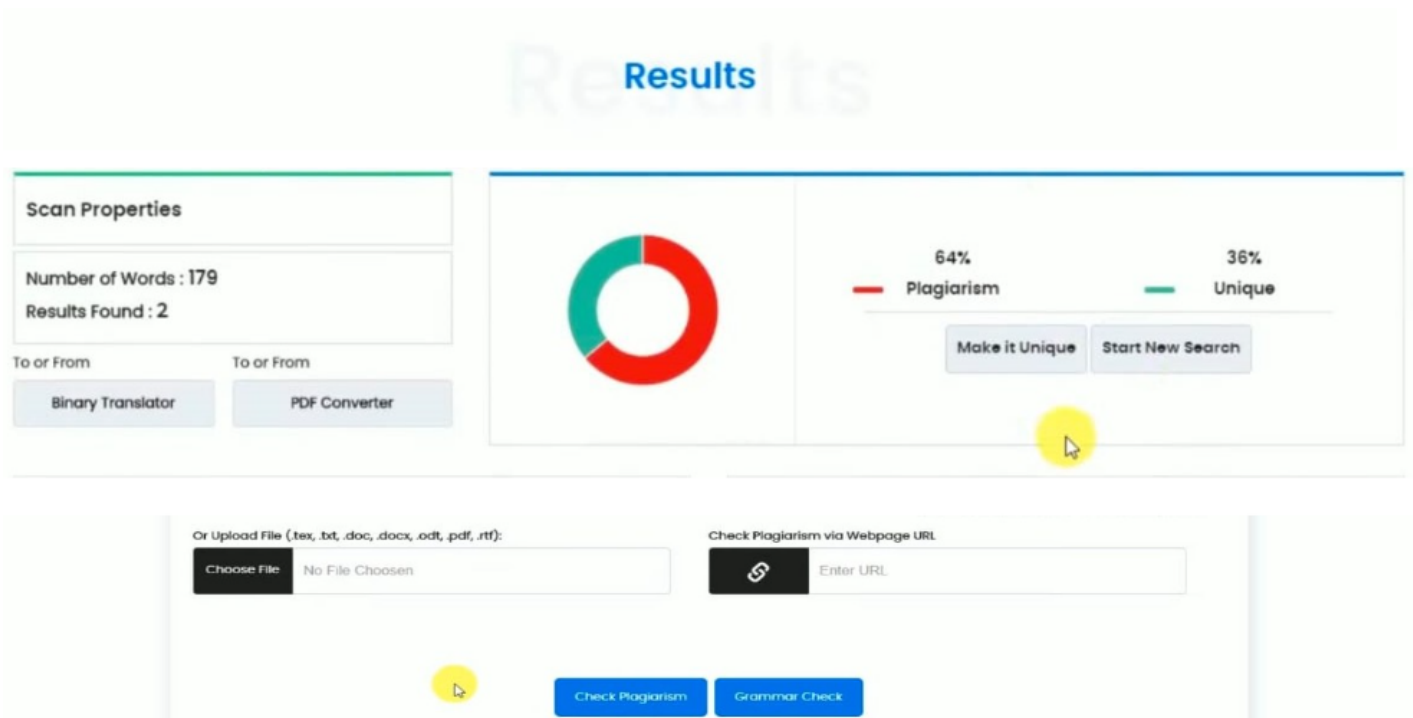
A product development canvas is the ground where the best possible ideas after the ideation session are pitched and nurtured to develop. The Product Canvas is a strategic product planning tool that allows us to quickly capture, describe, challenge, and pivot your product strategy on just a single page.

 <b>Purpose</b> What is the purpose of this concept you're developing? Does it solve a problem, or is it enhance a certain experience? Is it serving a need or it is trying to create a new need or tap an untapped need?	 <b>Product Experience</b> Define what your customer should feel like when he uses your product/service? What emotions, feelings, would define his experience? Feeling of comfort, convenience, or feeling of buying more with less cost conscious or feeling of greater security/safety etc.	 <b>Customer Revalidation</b> Once you're finished with your feature set, test with the customer/user if the features, functions are useful. Speak to the customer/user.
<b>To find plagiarism in text</b> <b>To find plagiarism in images</b>	 <b>Product Functions</b> Functions are a products answer to user problems/needs. They do something that user wants. They are often verbs in nature. Every function is powered by many features. Multitasking is a function. Browser tabs is a feature that powers the multitasking feature. A function can have one or more features powering it. Functions are very generic in nature, features are often more specific. Functions can be similar to product experience: safety (product function) provides a feeling of safety (product experience). <b>Finds Images difference</b> <b>Finds plagiarism in text and images</b> <b>Finds % of same text</b>	<b>Needs GUI</b> <b>Fast and Efficient</b> <b>provides security</b> <b>perfect to find plagiarism</b> <b>needs video format</b>
 <b>People</b> Who is the key customer segment who will use this product /service or the end product of the concept you're pursuing? Write here about them, describe them a little.	 <b>Product Features</b> Product features are specific. One or more features will power a function. Anti-lock Brakes, Airbags are features that power the safety function. Browser tabs, Apple's home button to multitask between apps are features powering the multitasking function. Each feature will have many components/sub components powering it. Sometimes a very regular component becomes a feature in itself. Like car doors is a major component and a feature at the same time powering the in car entertainment function powering entertainment as a product experience.	 <b>Reject, Redesign, Retain</b> Post customer validation, reject those functions or features that the customers didn't find useful. Redesign those that were partially useful and retain those that the user benefit with this unit all functions/features are accepted.
<b>Content Creator</b> <b>Students</b> <b>Teachers</b> <b>Visitors</b>	 <b>Components</b> Components build up the features. For a airbag it will comprise a lot of components like bags, triggers etc. that go into making it. For a tabbed browser it will comprise of various chunks of code that will make the tabs work. In cases where the feature is a major component, you could list here the auxiliary components that are required to make the major component work. You can also list new adjustments and innovations you're planning here at the component level. <ul style="list-style-type: none"> <li>- Computer</li> <li>- Pycharm</li> <li>- difflib</li> <li>- sequence Matcher</li> <li>- numpy</li> <li>- cv2</li> <li>- python 3,8 or higher</li> <li>- html 5</li> </ul>	<b>To merge it with legal party</b> <b>only text plagiarism</b> <b>No support for mp4</b> <b>only pdf option</b>

[3.4 Product Development Canvas]

### 3.5 Rough Prototype

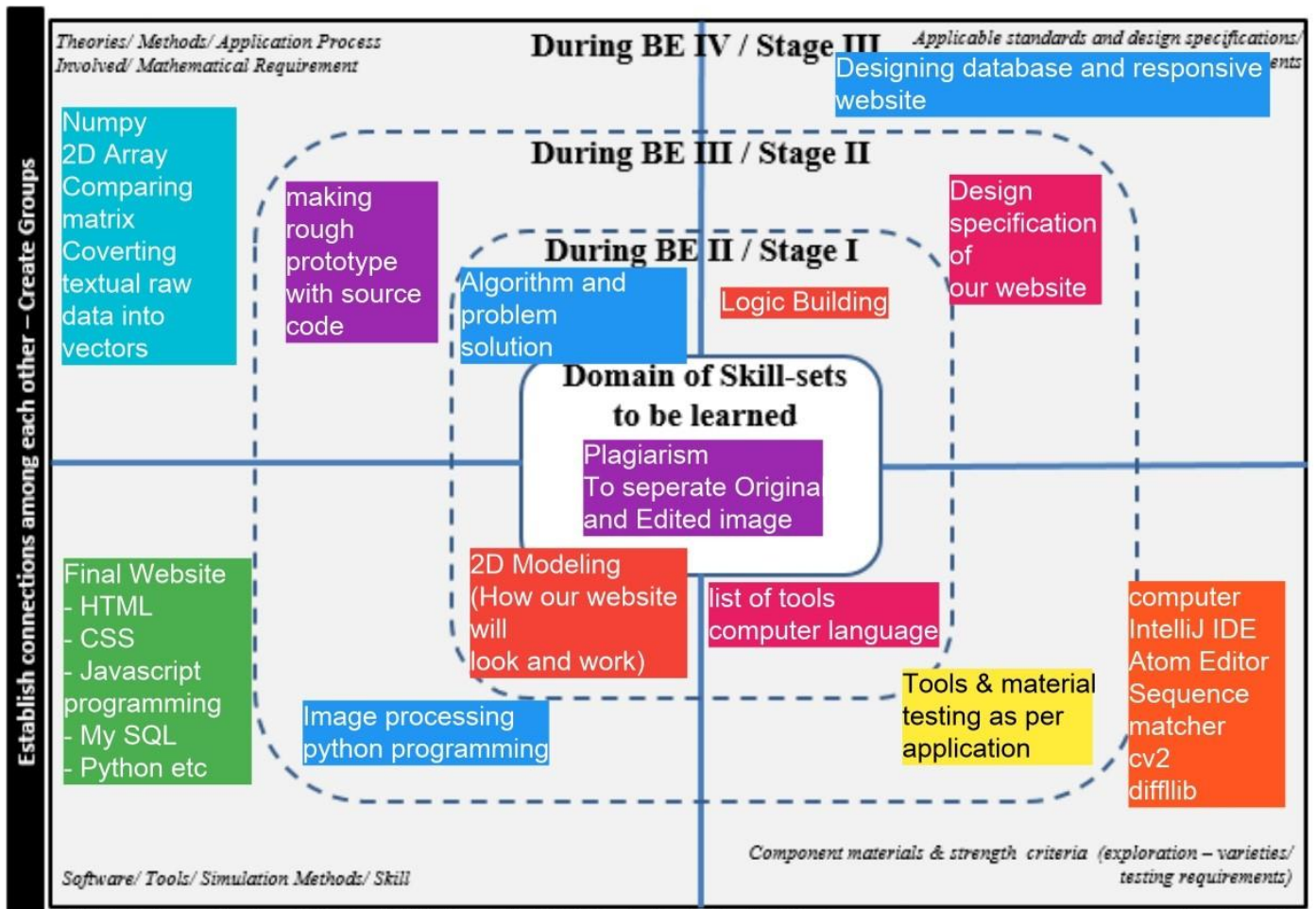
An early sample or model built to test a concept or process. The prototype is bound to have loose wires and rough edges.



[3.5 Rough mock ups for the implementation]

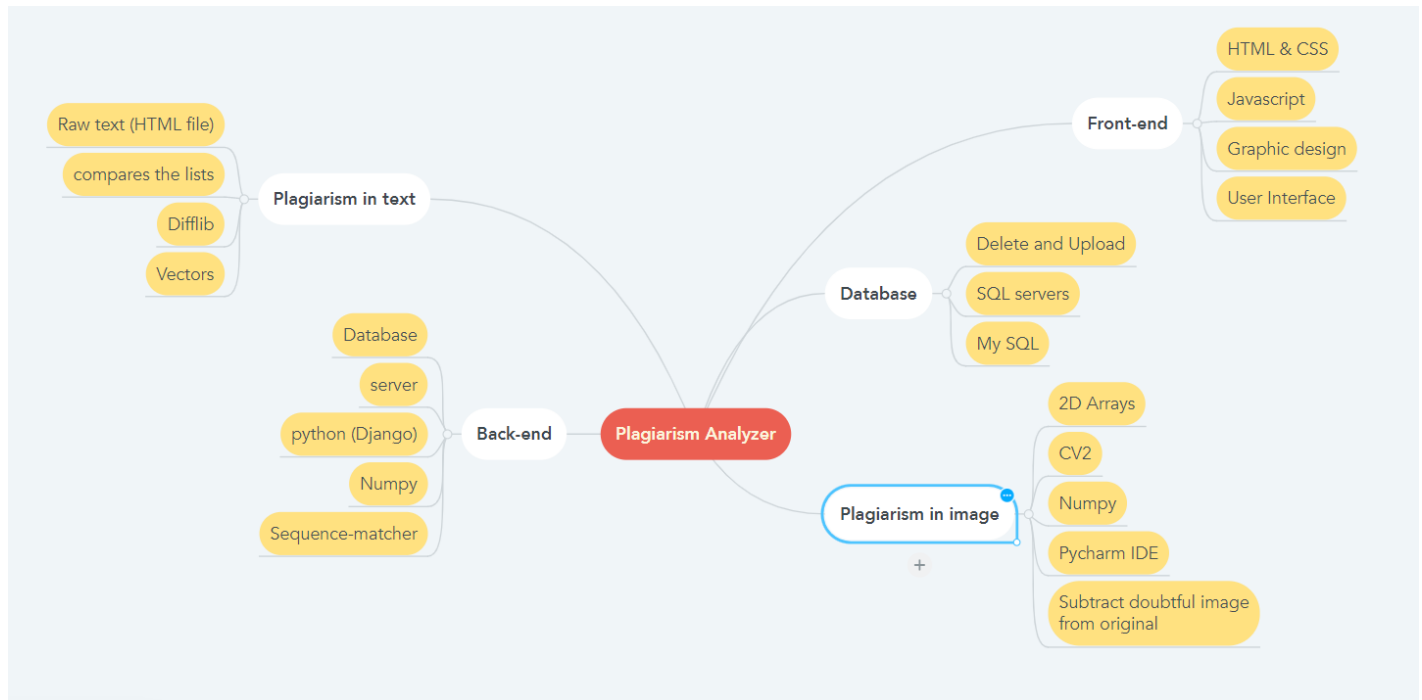


### 3.6 Learning Need Matrix:



[3.6 LNM]

### 3.7 Mind Map:



[3.7 Mind Map]

## **Chapter 4**

### **Conclusion & Future Work**

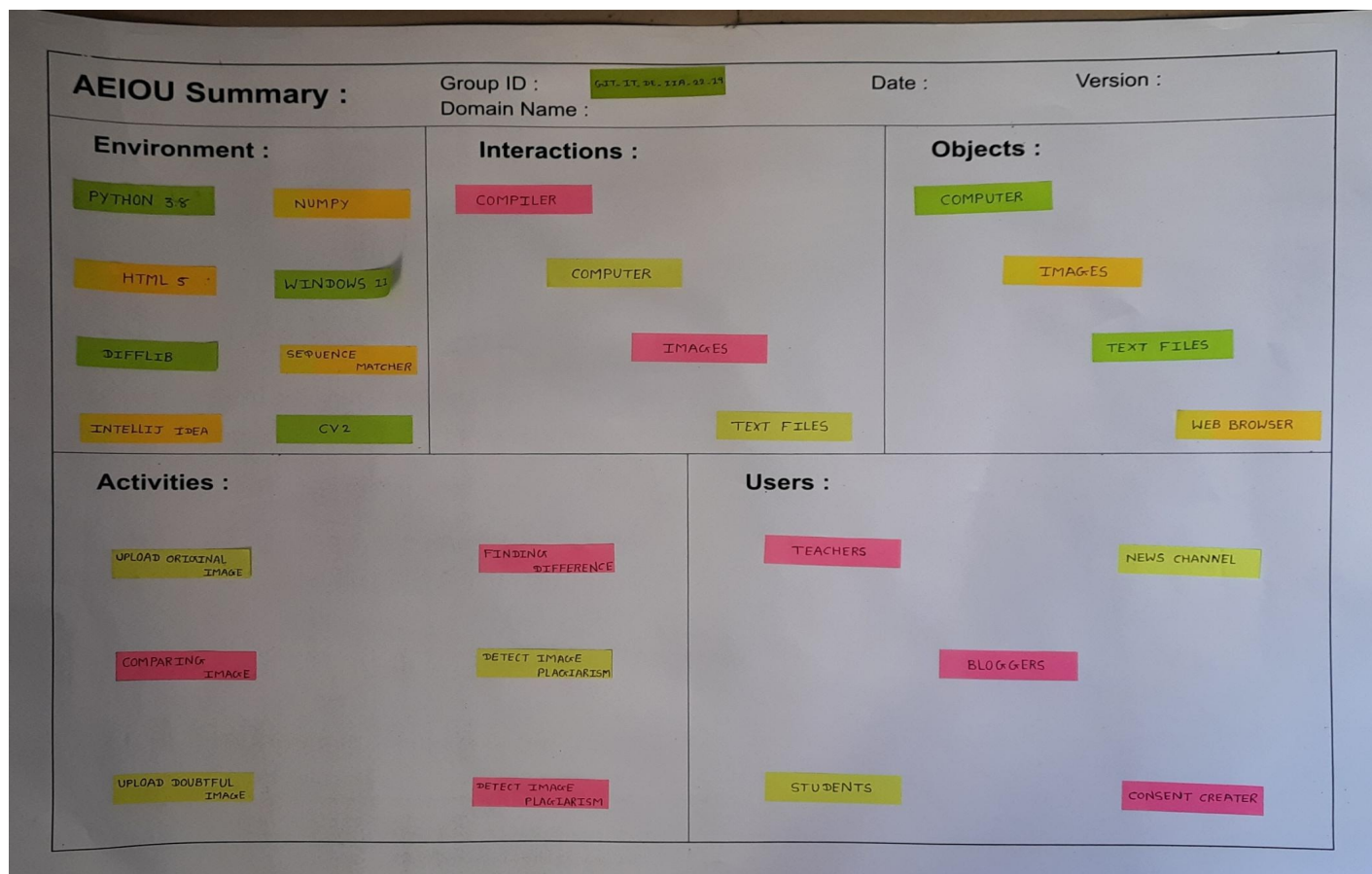
#### **4.1 Conclusion**

This is to conclude that the project assigned to us was worked upon with a sincere effort. This will definitely be some useful in students and even working professional's life and I'd like to examine three sectors that could benefit from a similar plagiarism checker product – art, music, and video. All of these industries currently have a number of outstanding legal cases and disputes over plagiarized content. With the right technology, plagiarism in all of these fields could be minimized through detection and checking prior to distribution.

## Appendix A

### Photos of Canvases

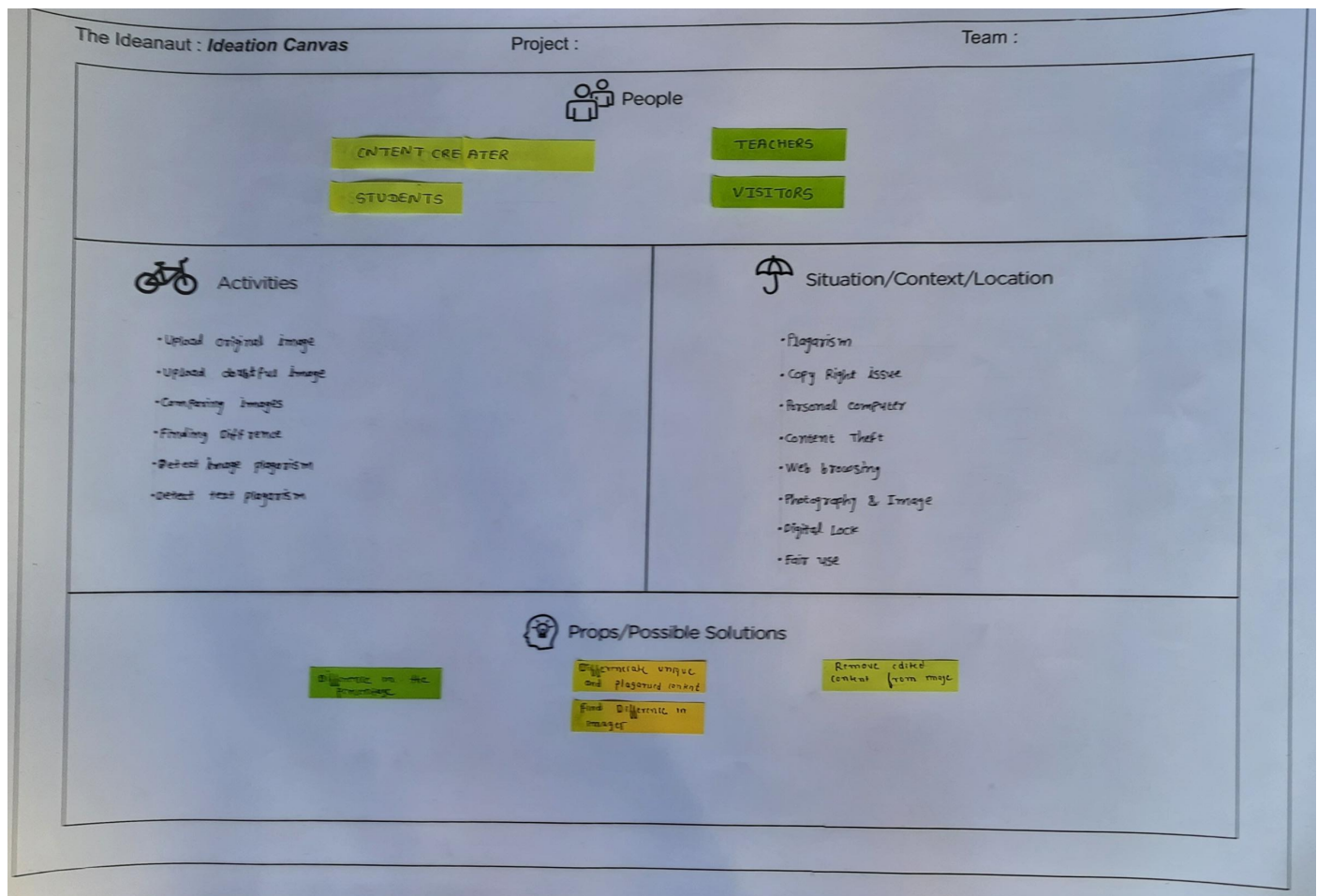
#### AEIOU Canvas:



## Empathy Canvas:

<b>Design For</b> <span style="background-color: yellow;">GIT-IT-DE-11A-22-IT</span>		<b>Design By</b>	
<b>Date</b>		<b>Version</b>	
<b>USER</b>		<b>STAKEHOLDERS</b>	
<span style="background-color: yellow;">TEACHERS</span> <span style="background-color: yellow;">NEWS CHANNELS</span>		<span style="background-color: pink;">DEVELOPERS</span>	
<span style="background-color: yellow;">CONSENT CREATOR</span>			
<span style="background-color: yellow;">STUDENTS</span> <span style="background-color: yellow;">BLOGGERS</span>			
<b>ACTIVITIES</b>			
<span style="background-color: pink;">UPLOAD ORIGINAL IMAGE</span>		<span style="background-color: yellow;">FINDING DIFFERENCE</span>	
		<span style="background-color: pink;">DETECT IMAGE PLAGIARISM</span>	
<span style="background-color: yellow;">UPLOAD DOUBTFUL IMAGE</span>		<span style="background-color: pink;">COMPARING IMAGES</span>	
		<span style="background-color: yellow;">DETECT TEXT PLAGIARISM</span>	
<b>STORY BOARDING</b>			
<b>HAPPY</b> It make edits and NFT images, it gets stolen pretty quick which results into loss, with the help of it we can prevent from getting stolen.			
<b>HAPPY</b> I occasionally uploads blog on the internet and i have seen my content on somones else's website, claimed as their own's, with the help of this i can finally prove its my original work and can get their content removed.			
<b>SAD</b> It is very good project but it only detects the plagiarism not remove it.			
<b>SAD</b> It can not find if the video is plagiarised or not.			

## Ideation Canvas:

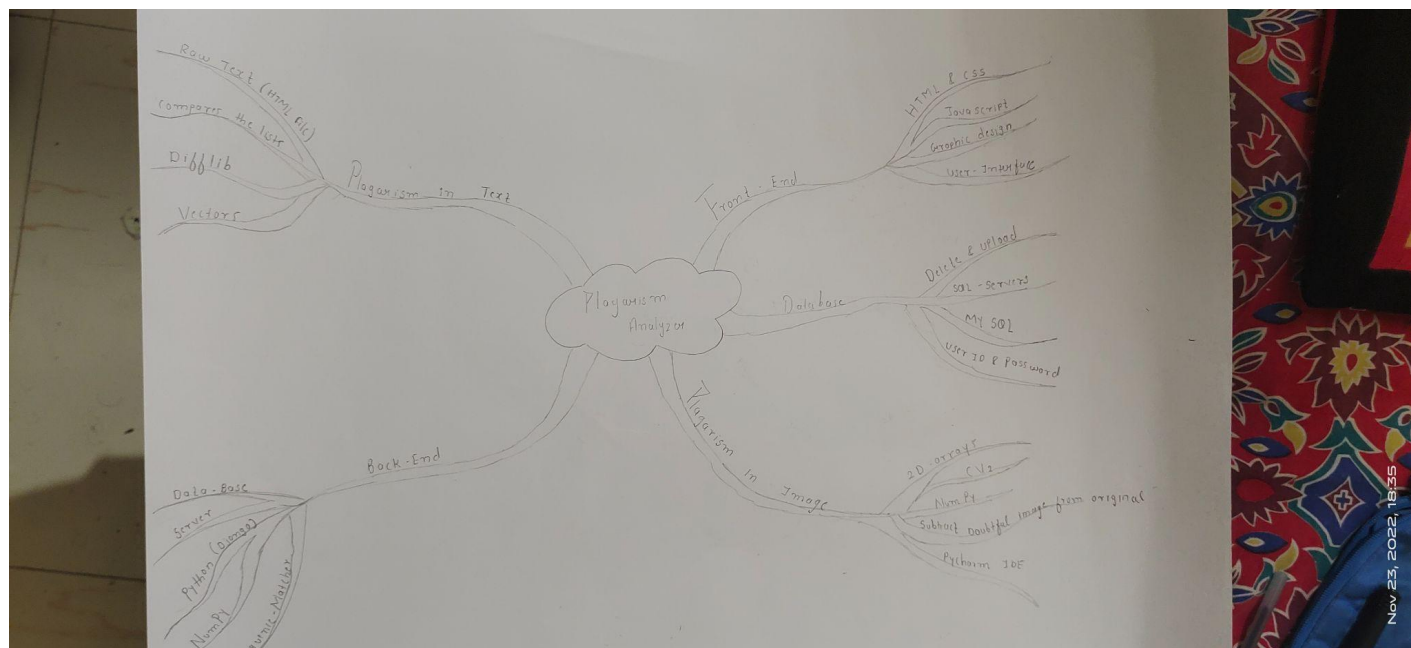




# Product Development Canvas:

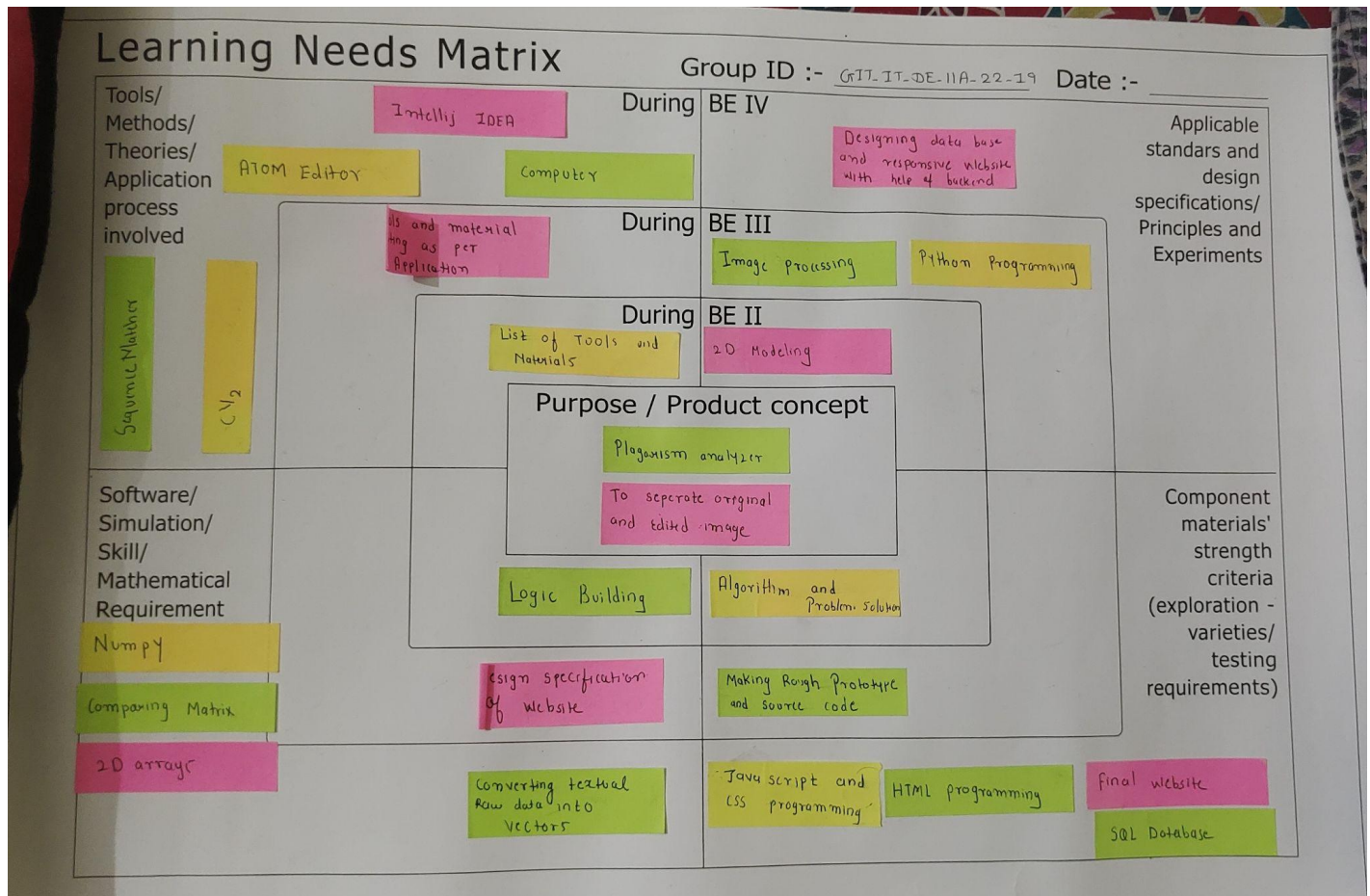
Product Development Canvas		Team/Date/Version :
<p><b>🔍 Purpose</b></p> <p>What is the purpose of this concept you're developing ? Does it solve a problem, or it enhances a certain experience ? Is it serving a need or it is trying to create a new need or tap an untapped need ?</p> <p>TO FIND PLAGIARISM IN TEXT</p> <p>TO FIND PLAGIARISM IN IMAGE</p>	<p><b>👑 Product Experience</b></p> <p>Define what your customer should feel like when he uses your product/service ? What emotions, feelings would define his experience ? Feeling of comfort, convenience, or feeling of buying more with less (cost conscious) or feeling of greater security, safety etc.</p> <p><b>🔧 Product Functions</b></p> <p>Multitasking is a function. Browser tabs is a feature that powers the multitasking feature. A function can have one or more features powering it. Functions are very generic in nature, features are often more specific. Functions can be similar to product experience. Safety (product function) provides a feeling of safety (product experience).</p> <p>FIND PLAGIARISM IN TEXT</p> <p>FIND JOB SAME TEXT</p> <p><b>📦 Product Features</b></p> <p>Product features are specific. One or more features will power a function. Antilock Brakes, Airbags are features that power the safety function. Browser tabs, Apple's home button to multitask between apps are features powering the multitasking function. Each feature will have many components/sub components powering it. Sometimes a very popular component becomes a feature itself. Like car stereo is a major components and a feature at the same time powering the in car entertainment function powering entertainment as a product experience.</p> <p>USES LATEST AI</p> <p>PROVIDES SECURITY</p> <p>FIND TEXT PLAGIARISM</p> <p><b>⚙️ Components</b></p> <p>Components build up the features. For a airbag it will comprise a list of component like bags, triggers etc. that go into making it. For a tabbed browser it will comprise of various chunks of code that will make the tabs work. In cases where the feature is a major component, you could list here the auxiliary components that are required to make the major component work. You can also list new adjustments and innovations you're planning here at the component level.</p> <p>COMPUTER</p> <p>NUMPY</p> <p>PYCHARM</p> <p>CV2</p> <p>SEQUENCE MATCHER</p> <p>PYTHON 3.6</p> <p>DIFF LIB</p> <p>HTML 5</p>	<p><b>📋 Customer Revalidation</b></p> <p>Once you're finished with your feature set, test with the customer / user if the features, functions are useful. Speak to the customer / user.</p> <p>Needs GUI</p> <p>Fast and Efficient</p> <p>Provider Security</p> <p>Perfect to find Plagiarism</p> <p>Needs Video Format</p> <p><b>🧪 Reject, Redesign, Retain</b></p> <p>Post customer validation, reject those functions or features that the customers didn't find useful. Redesign those that were partially useful and retain those that met the bar. Iterate with this until all functions/features are accepted.</p> <p>To merge it with legal panel</p> <p>Only Text Plagiarism</p> <p>Needs MP4 Format</p> <p>No need of PDF</p>
<p><b>👤 People</b></p> <p>Who is the key customer segment who will use this product /service or the end product of the concept you're putting ? Write here about them, describe them a little.</p> <p>Teachers</p> <p>Content Creator</p> <p>Visitors</p> <p>Visitors</p> <p>Students</p>		

## Mind Map:





## Learning Need Matrix:



## REFERENCES

---

- [1] [www.google.com](http://www.google.com)
- [2] [www.youtube.com](http://www.youtube.com)
- [3] [www.duplichecker.com](http://www.duplichecker.com)
- [4] [www.scribbr.com](http://www.scribbr.com)